Puerto Rico Address Data Working Group
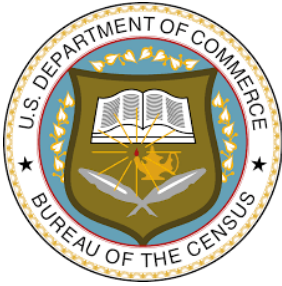
April 2, 2019

Kick-Off Meeting
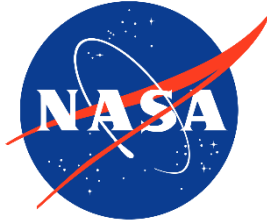
United States® Census 2020

FGDC.GOV
FEDERAL GEOGRAPHIC DATA COMMITTEE

Address Subcommittee

Agenda

Welcome and Introductions

Background

Working Group Objectives

Building a Proven Practices Inventory

Next Steps

United States® Census 2020

Welcome
Participating Agencies and Organizations

Nancy Potok, Ph.D.

Chief Statistician of the United States

Chief of Statistical and Science Policy at the

Office of Management and Budget

Agenda

Welcome and Introductions

Background

Working Group Objectives

Building a Proven Practices Inventory

Next Steps

United States® Census 2020

Why this Working Group?

Various intergovernmental entities have identified the need for an inclusive federal government conversation about the collection and management of Puerto Rico addresses.

Federal Proven Practices for Puerto Rico Addresses

Subcommittee on Disaster Reduction

Address Subcommittee

October 2018

Puerto Rico Address Data Workshop

Whether fulfilling the mission to distribute aid after a disaster, enumerate and tabulate the population, or deliver mail on-time, many government entities have a stake in managing PR address data in order to meet their respective mission objectives

Agenda

■ Welcome and Introductions

■ Background

■ **Working Group Objectives**

■ Building a Proven Practices Inventory

■ Next Steps

United States®
Census
2020

Purpose Statement

Provide a structured approach to identifying, documenting, and socializing a collection of federal Proven Practices in Puerto Rico address data management:

Promotes consistency, transparency, and strategic prioritization in the federal address collection and management process;

Documents agency missions and use cases; and

Aligns with the goals of the FGDC Address Theme and Subcommittee to contribute to the National Geospatial Data Infrastructure (NSDI).

United States® Census 2020

Working Group Objectives

The objectives below were noted in the October 2018 Puerto Rico Address Data Workshop Minutes:

Document and socialize **current Proven Practices** for collecting, managing, and sharing Puerto Rico address data within federal agencies
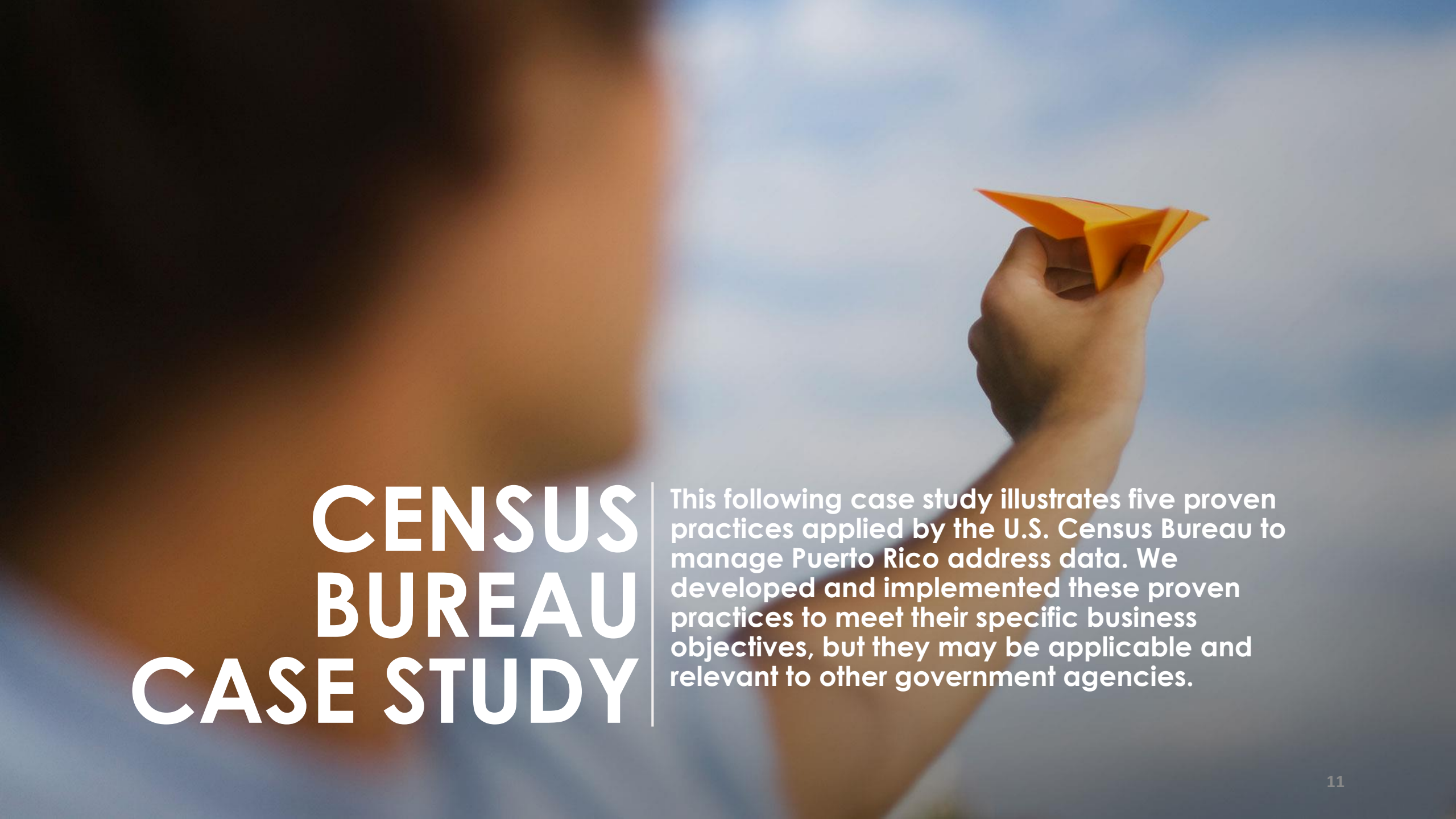
Encourage the implementation and update of existing national address standards for Puerto Rico addresses managed by federal agencies

Publish recommendations for address content that corresponds to the Proven Practices gathered across federal agencies

United States® Census 2020

Agenda

Welcome and Introductions

Background

Working Group Objectives

Building a Proven Practices Inventory

Next Steps

United States® Census 2020

CENSUS BUREAU CASE STUDY

This following case study illustrates five proven practices applied by the U.S. Census Bureau to manage Puerto Rico address data. We developed and implemented these proven practices to meet their specific business objectives, but they may be applicable and relevant to other government agencies.

Puerto Rico Address Types

The Census Bureau maintains five (5) different varieties of Puerto Rico addresses in our internal address inventory, called the Master Address File, or MAF. Each of these addresses is residential in nature, either for a single or multi-family structure, an institutional or group quarters facility, or military housing.

General

1234 Calle Juan Doe
San Juan, PR 00926
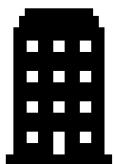
Apartment Complex

8 Vista Suites APT 8910
San Juan, PR 00926

Other

Highway Contract
Rural Route
P.O. Box

Urbanización

Urb Smith
567 Calle A
San Juan, PR 00926

Area Name

Barrio Juan Sectore Smith
Carretera 12
San Juan, PR 00926

NOTE: These examples are fictional addresses that are not protected from disclosure by the Data Stewardship and Confidentiality requirements of 13 U.S.C. §9.

Majority Address Type by Block

Inset B

Inset A

5 Miles

A - Ponce Vicinity

1 Mile

B - San Juan Vicinity

1 Mile

		Apartment Complex
🏢	🟥	Apartment Complex
🏠	🟩	Area Name
✉️	🟧	General
🏢🏢	🟪	Urbanización
🔍	🟨	Other
	⬜	No Addresses

United States® Census 2020

13
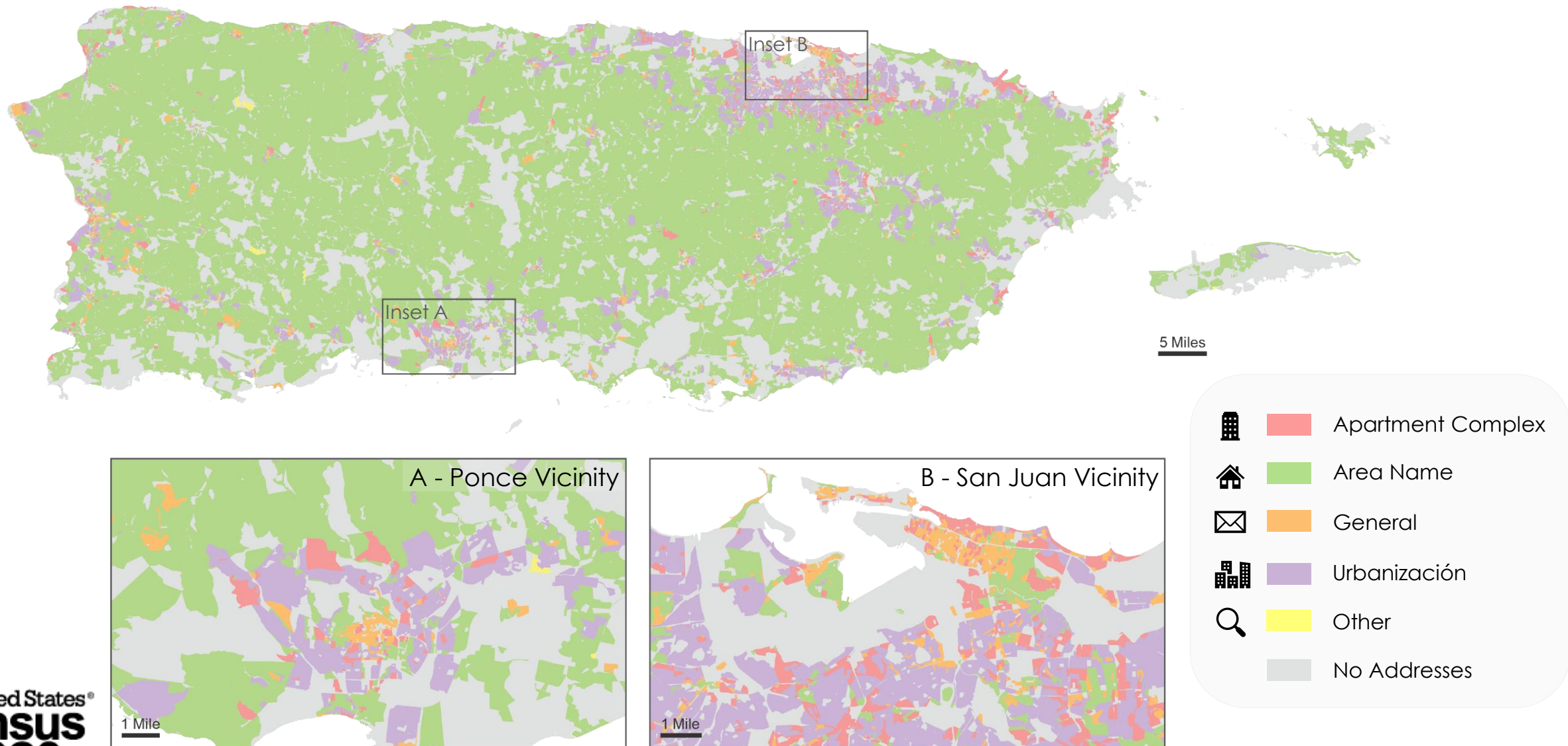
Puerto Rico Address Profile - Census

The largest categories of address types in the Puerto Rico Address inventory – *Urbanización* and *Area Name* – illustrates the fact that the majority of Puerto Rico addresses do not match the conventional General format, which makes it difficult for systems to match data unless they have some **customizations** in place.

Other
4%

General
11%

Area Name
37%

Urbanización
35%

Apartment Complex
13%

United States®
Census
2020
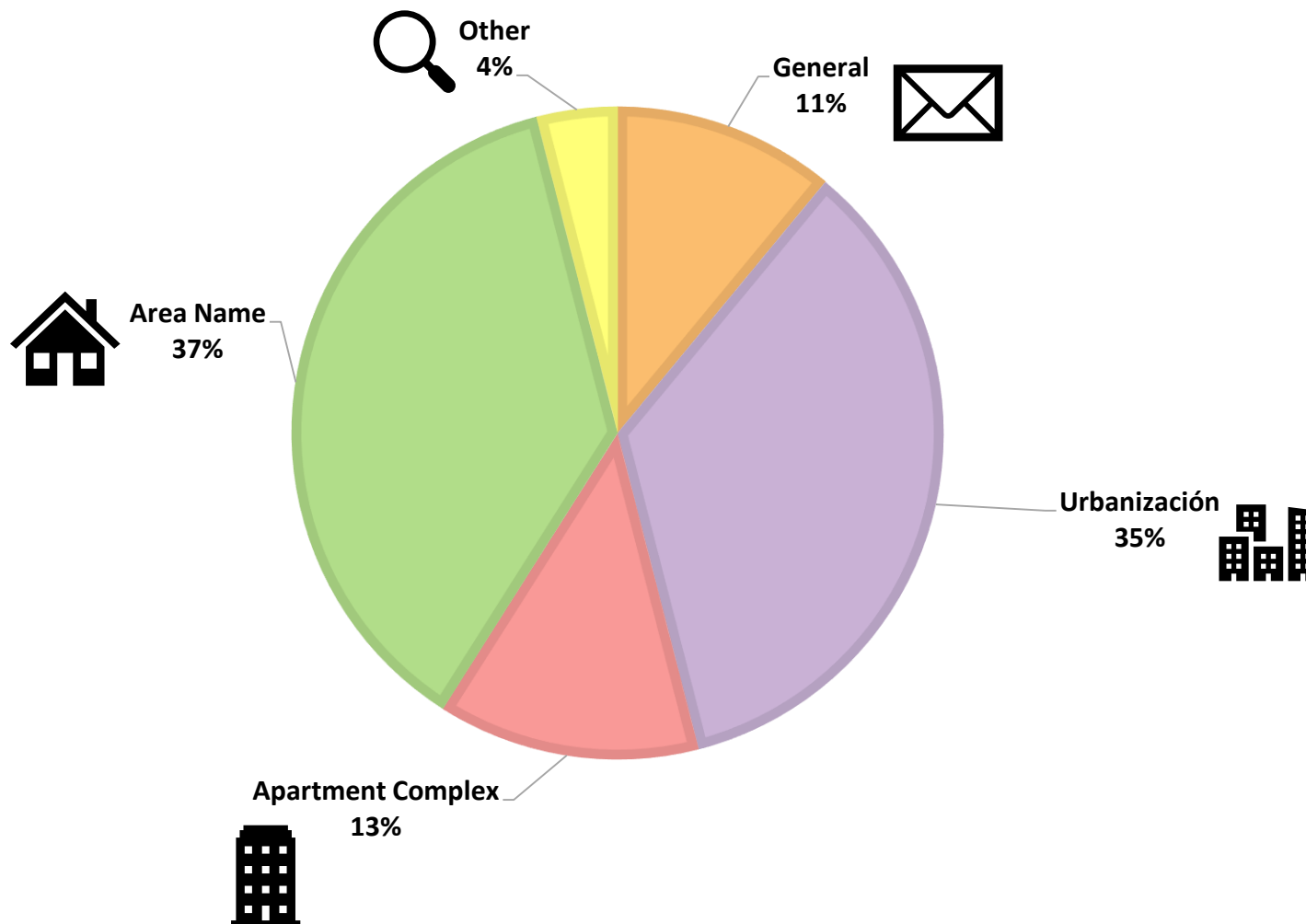
Address Management Proven Practices

Standardize the Data

Clean the Data

Match Data

Maintain Data Source and History

Gather Data Regularly

Standardize the Data

To account for various fields and words used in Puerto Rico addresses, the Census Bureau utilizes **standardized naming conventions** to capture and organize data.

Standardize the Data

To account for the variety of fields and words used to reference Puerto Rico addresses, the Census Bureau uses industry standard codes that follow U.S. Postal Service (Publication 28) addressing standards.

ADDRESS ELEMENT	SOURCE VERSION	STANDARDIZED VERSION
Street Name	Calle 1	CLL 1
Unit Descriptor	Apartment	APT
Building Descriptor	Edificio	EDIF
Urbanización	Urbanización	URB
Apartment Complex Name	Residenciál	RES
Area Name 1/Area Name 2	Barrio	BO
KMHM	KM 01.5	1.5
Non-City Style	Rural Route 3	RR 3

Clean the Data

Recognizing and rearranging **different fields and formats** allows software to recognize more variety in how Puerto Rico addresses are represented.

Clean the Data

In addition to standardizing language, the Census Bureau developed a method that enables it to "clean up" address data formatting and sequencing, enabling analysis using the whole address or individual address components.

INCOMING DATA

House Number	Urbanización & Street Name	Unit Number	ZIP Code

CENSUS FORMAT

House Number	Urbanización	Street Name	Unit Number	ZIP Code	Area Name

WHOLE ADDRESS ANALYSIS

House Number	Urbanización	Street Name	Unit Number	ZIP Code

Algorithm compares individual components; creates the whole address; and then performs an equivocated match

Match Data

The addition of a **whole address matching solution** bolsters the typical 'exact' and 'equivocated' match formulas that are typically used on incoming data.

Match Data

The Census Bureau match process uses both individual address components as well as the **whole address.**

Incoming Address: **6 INDUSTRIAL WAY**

1st attempt

Seeks an **exact match** of data

Existing Census Address: **6 INDUSTRIAL WAY**

match

Finds an exact match on all address components during 1st attempt

NOTE: This example is a fictional address that is not protected from disclosure by the Data Stewardship and Confidentiality requirements of 13 U.S.C. §9.

Match Data

The Census Bureau match process uses both individual address components as well as the **whole address.**

Incoming Address: **126 <mark>CALLE</mark> DR RAMONA BETA N**

1st attempt

Seeks an **exact match** of data

2nd attempt

Equivocated match comparing **individual components**

Existing Census Address: **126 <mark>CLL</mark> DR RAMONA BETA N**

no match

Fails an exact match

match

Equivocated match on street (CALLE vs. CLL)

NOTE: This example is a fictional address that is not protected from disclosure by the Data Stewardship and Confidentiality requirements of 13 U.S.C. §9.

Match Data

The Census Bureau match process uses both individual address components as well as the **whole address.**

Incoming Address : **3W20 CALLE 40 URB BUCARABONES**

1st attempt	2nd attempt	3rd attempt
Seeks an **exact match** of data	**Equivocated** match comparing **individual components**	Equivocated match using **whole address** and **Census string comparator**

Existing Census Address: **3W20 CLL 40 URB BUCABABONES**

no match	no match	match
Fails an exact match	*Fails an equivocated match of individual address components*	*Mathematically these addresses are the same; an equivocated match*

NOTE: This example is a fictional address that is not protected from disclosure by the Data Stewardship and Confidentiality requirements of 13 U.S.C. §9.

Match Data

The Census Bureau match process uses both individual address components as well as the **whole address.**

Incoming Address: **300 <mark>PASEO</mark> CASABLANCA APT <mark>16</mark>**

1st attempt	2nd attempt	3rd attempt	4th attempt
Seeks an **exact match** of data	**Equivocated** match comparing **individual components**	Equivocated match using **whole address** and **Census string comparator**	**Equivocated** match using **whole address** and **Jaro-Winkler** algorithm

Existing Census Address: **300 <mark>PSO</mark> CASABLANCA APT <mark>16 COND PASEO CASABLANCA</mark>**

no match	no match	no match	match
Fails an exact match	*Fails an equivocated match of individual address components*	*Fails to mathematically establish an equivocated match*	*Mathematically, these addresses are the same; an equivocated match*

In Practice: Standardize, Clean and Match Data

A combination of the standardization, cleaning and matching data steps

Phase	Example 1	Example 2
Incoming Address	Calle Main 11A Urbanizacion Coastal Gardens ↓	Condominio San Juan Garden Edificio A Apart 100 ↓
Standardized Address	CLL MAIN **11A** URB COASTAL **GARDENS** ↓	COND SAN JUAN **GARDEN** **EDIF** A **APT** 100 ↓
Equivocated Match	CLL MAIN **A11** URB COASTAL **GRDNS**	COND SAN JUAN **GRDN** **BLDG** A **UNIT** 100

NOTE: These examples are fictional addresses that are not protected from disclosure by the Data Stewardship and Confidentiality requirements of 13 U.S.C. §9.
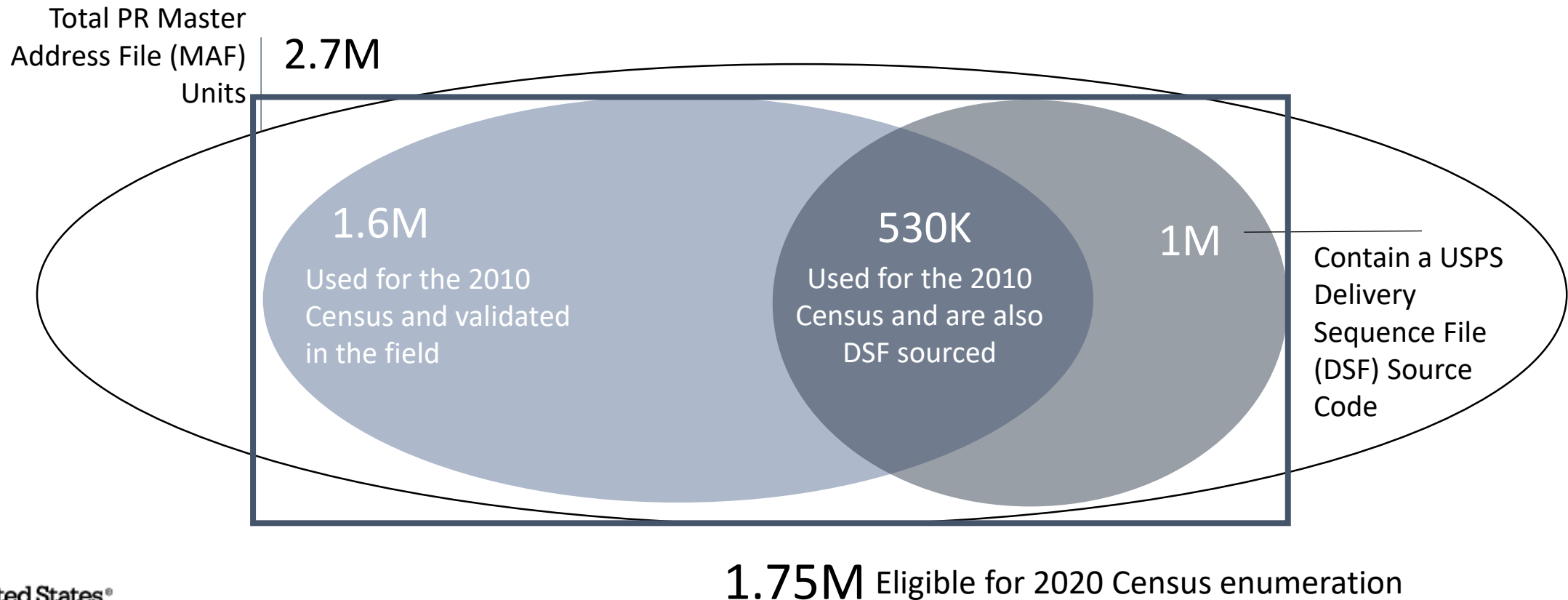
Maintain Data Source and History

Source and history information enables the Census Bureau to **develop a longitudinal pattern** that can be compared to and validated by other authoritative address sources.

United States®
Census
2020

Maintain Data Source and History

This proven practice enables the Census Bureau to establish a longitudinal pattern that can be compared to other sources and helps validate that an address should be included in the Decennial Census and other surveys.

Total PR Master Address File (MAF) Units

2.7M

1.6M
Used for the 2010 Census and validated in the field

530K
Used for the 2010 Census and are also DSF sourced

1M

Contain a USPS Delivery Sequence File (DSF) Source Code

1.75M Eligible for 2020 Census enumeration

United States® Census 2020

The more successful sources you can attach to an address, the higher the confidence that it's an accurate address and should be included in Census enumeration efforts

Gather Data Regularly

Collecting data regularly allows the Census Bureau to maintain records that **capture change in addresses**, enabling the agency to manage information requests and deliver on mission objectives

United States®
Census
2020

MAF Puerto Rico Address Sources

The Census Bureau's Master Address File (MAF) is built from multiple sources and partners in an attempt to maintain a current, comprehensive inventory of accurate data.

United States Postal Service

- Delivery Sequence File (DSF)
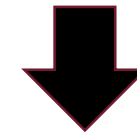- Locatable Address Conversion system (LACS)

Commonwealth of Puerto Rico

- Geographic Support System (GSS) Partnership Program
- Local Update of Census Addresses (LUCA)

Census Field Operations
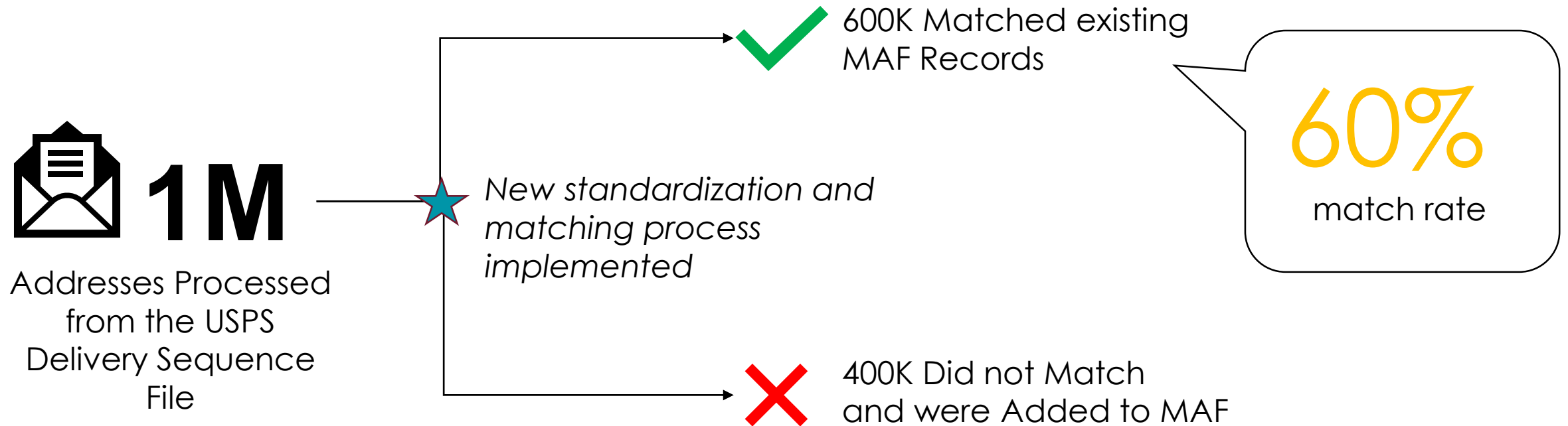
- 2000 Decennial Census
- 2010 Decennial Census

CENSUS BUREAU MASTER ADDRESS FILE
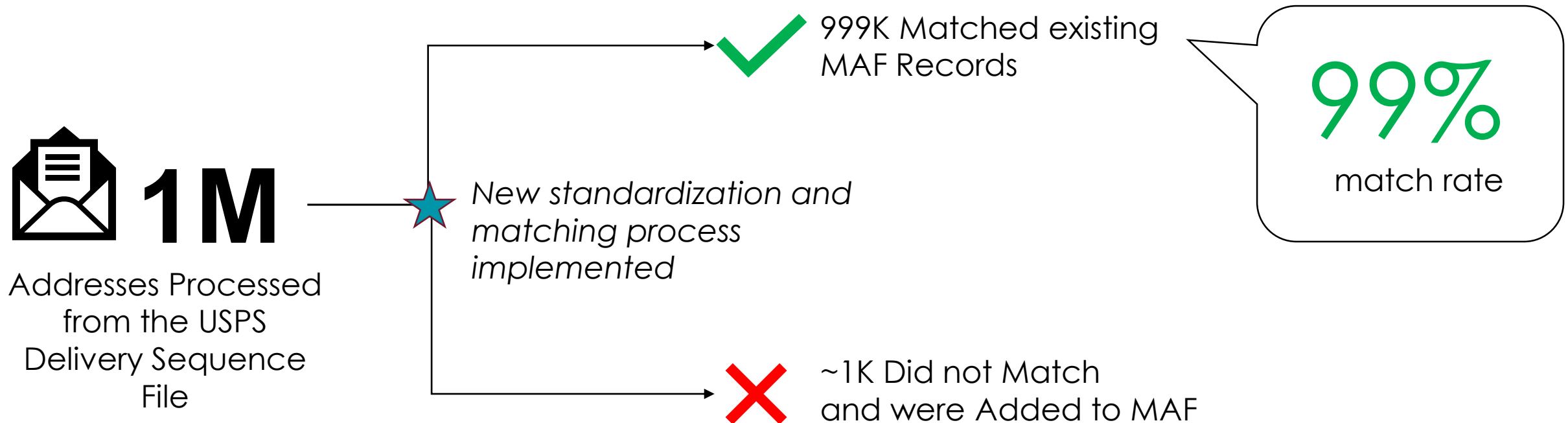
Gather Data Regularly

The Census Bureau applied our new and upgraded standardization and matching processes to the USPS Delivery Sequence File for the first time in Spring 2016; this initial comparison resulted in matches for nearly 60% of the addresses received.

1M

Addresses Processed from the USPS Delivery Sequence File

New standardization and matching process implemented

✓ 600K Matched existing MAF Records

✗ 400K Did not Match and were Added to MAF

60%
match rate

Gather Data Regularly

After five (5) "refreshes" of the MAF with the Delivery Sequence File, the rate of new addresses has stabilized at approximately 1% every six (6) months. Our upgraded processes enable the Census Bureau to track change and build data sets with a high level of confidence.

1M

Addresses Processed from the USPS Delivery Sequence File

New standardization and matching process implemented

✔ 999K Matched existing MAF Records

✘ ~1K Did not Match and were Added to MAF

99% match rate

Agenda

■ Welcome and Introductions

■ Background

■ Working Group Objectives

■ Building a Best Practices Inventory

■ **Next Steps**

United States® Census 2020

Next Steps

- Participate in **subsequent working group** phone calls and meetings (bi-weekly)

- Finalize the **PR Working Group Charter**

- Collaborate with your respective teams to identify and **document proven practices inventory** from your agencies

- **Build an arsenal of proven practices** for dissemination to the FGDC Address Subcommittee in support of intergovernmental mission objectives (writing to begin Mid-May)

- **Save the Date!** Week of June 24th – next in-person meeting
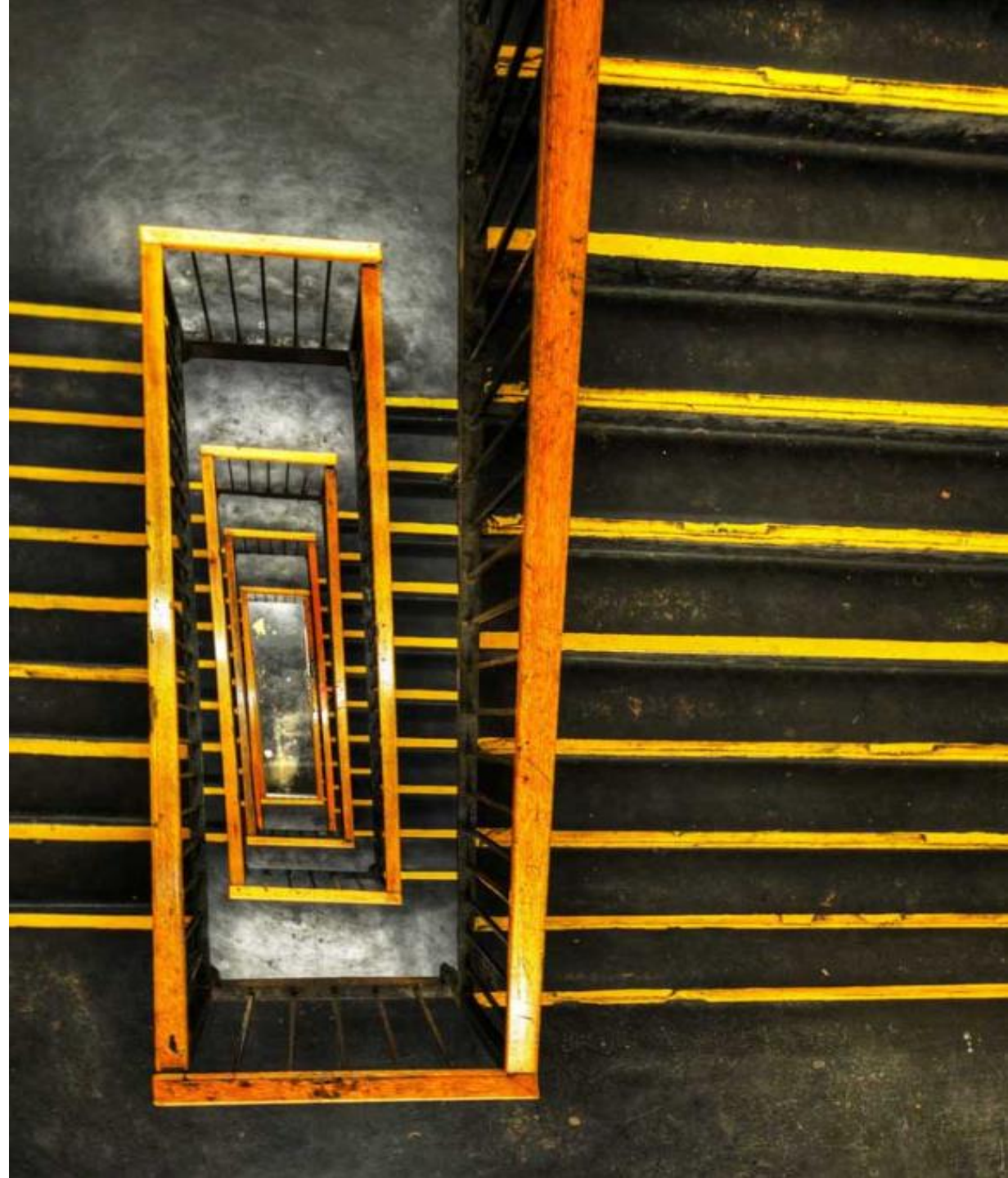
Contact Information

For more information about the PR Address Data Working Group:

Gregory Hanks Jr.
Census Bureau
Executive Champion, PRADWG
Phone: 301-763-3093
Gregory.F.Hanks.Jr@census.gov

Karla Kaywork
Census Bureau
Contact Coordinator, PRADWG
Phone: 301-763-7633
Karla.C.Kaywork@census.gov

For more information about the FGDC Address Subcommittee:

Lynda Liptrap
Census Bureau
Co-Chair, FGDC Address Subcommittee
Phone: 301-763-1058
Lynda.A.Liptrap@census.gov

Steve Lewis
Department of Transportation
Co-Chair, FGDC Address Subcommittee
Phone: (202) 366-9223
Steve.Lewis@dot.gov

For more information about the Census Bureau's management of Puerto Rico address data:

Stuart Irby
Census Bureau
Chief, Address Standards, Criteria, and Quality Branch
Phone: 301-763-9180
Stuart.C.Irby@census.gov

Laura Hogberg
Census Bureau
Team Leader, Address Standards, Criteria, and Quality Branch
Phone: 301-763-7254
Laura.Kathleen.Hogberg@census.gov

Puerto Rico Address Data Working Group

Thank You!